proteins determines the cell function. Since proteins are long and linear complex molecules, they "fold" to give a 3D shape. Biologists have identified four levels of structure which can influence the protein's function:

1. Primary structure -- the sequence of amino acids
2. Secondary structure -- the presence or absence of small "sub-folds". These are regular patterns formed by local folding of the protein (e.g., helices and sheets).
3. Tertiary structure -- the final 3D shape
4. Quaternary structure -- complexes formed with other proteins.

Please replace the paragraph at page 6, line 24 through page 7, line 2, with the following paragraph and Table:

Illustrated in Fig. 1 is a computer system embodying the present invention. A digital processor 13 executes invention software program 15 in working memory. The invention software program 15 receives as input 11 a subject amino acid (i.e., protein or DNA) sequence or subsequence. The input sequence/subsequence 11 is a text string (consisting of A's, C's, T's, and G's) for representing the sequence of amino acids. Each amino acid can be represented by one or more characters, an example of which is given in Table 1.

Table 1

Amino Acid	3-Letter Code	1-Letter Code
Alanine	Ala	A
Cysteine	Cys	C
Aspartate	Asp	D
Glutamate	Glu	E
Phenylalanine	Phe	F
Glycine	Gly	G

Histidine	His	H
Isoleucine	Ile	I
Lysine	Lys	K
Leucine	Leu	L
Methionine	Met	M
Asparagine	Asn	N
Proline	Pro	P
Glutamine	Gln	Q
Arginine	Arg	R
Serine	Ser	S
Threonine	The	T
Valine	Val	V
Tryptophan	Trp	W
Tyrosine	Tyr	X

Please replace the paragraph at page 7, lines 3 through 10 with the following paragraph:

Different amino acid sequences have different length text string representations. Hence, the input sequences to invention program 15 are of varying lengths. Using a predefined set 17 of known biological fragments, the invention software program 15 performs a comparison routine 19 against the subject amino acid sequence input 11. The comparison routine 19 effectively transforms the traditional text representation of the subject amino acid sequence 11 into a fixed length vector 23. That is, the comparison routine 19 transforms the input sequences of varying length into respective same length (i.e., uniform length) feature vectors 23.

Please replace the paragraph at page 11, lines 2 through 13 with the following paragraph: